Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER Stochastic Processes and their Applications 115 (2005) 661–677
stochastic processes and their applications
www.elsevier.com/locate/spa

Partially exchangeable processes indexed by the vertices of a k-tree constructed via reinforcement ☆

Pietro Muliere[a], Piercesare Secchi[b], Stephen Walker[c,*]

[a] *Dipartimento di Metodi Quantitativi, Università L. Bocconi, Viale Isonzo 25, 20135 Milan, Italy*
[b] *Dipartimento di Matematica, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milan, Italy*
[c] *Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, Kent CT2 7NZ, UK*

Received 14 June 2004; received in revised form 27 October 2004; accepted 15 November 2004
Available online 9 December 2004

Abstract

We define a reinforced stochastic process of random variables indexed by the vertices of a k-tree and with values in a Polish space. The work presents a natural extension from an exchangeable to a partially exchangeable setting of previous work done by the authors.

Keywords: Balanced design; Beta blanket; Binary regression; Dirichlet blanket; Martingale; Partial exchangeability

1. Introduction

In this paper we introduce new stochastic processes indexed by the vertices of a k-tree. While these processes are of interest in their own right, there are statistical applications in Bayesian nonparametric problems. As such, the paper can be viewed

☆The research of S.G. Walker is financed by a UK EPSRC Advanced Research Fellowship. The work of P. Secchi and P. Muliere is partially supported by MIUR, Italy.
*Corresponding author. Tel.: +44 0 1227 823800; fax: +44 0 1227 827932.
E-mail address: S.G.Walker@kent.ac.uk (S. Walker).

0304-4149/$ - see front matter
doi:10.1016/j.spa.2004.11.004

as a natural extension of the papers of Muliere et al. [6,7] from exchangeable to regression (i.e. non-exchangeable) problems via consideration of partially exchangeable sequences. Reinforcement will again be the fundamental tool for constructing the processes.

As motivation for our binary process described in Section 2, consider the Bayesian nonparametric binary regression model discussed and studied by Diaconis and Freedman [5]. They assume a balanced design for a medical trial involving m patients, each patient being identified with a unique sequence $\varepsilon_1 \dots \varepsilon_n$ with $\varepsilon_i \in \{0, 1\}$. Hence, for a design of size n, there are $m = 2^n$ patients. A binary outcome is then observed from each of the m patients. A prior for such data involves the construction of (dependent) random probabilities for each of the 2^n possible sequences which is dimensionally coherent. Our model also provides an easy extension to non-binary regression where we construct dependent random distribution functions. The level of dependence between pairs of these functions depends on the closeness of covariates associated with each function.

In the next section we develop the framework for our processes: this involves the use of k-trees. In Section 3 we introduce a process on a k-tree which can be used as a Bayesian nonparametric prior for the binary regression model discussed earlier in this introduction; we will then return to this example in Section 4. Section 5 extends the process introduced in Section 3 to a general regression problem where the outcome for each co-variable is continuous. We examine a further idea in Section 6 and we conclude with a discussion in Section 7, where we point out the relation of the current paper to previous work.

2. The geometry of a *k*-tree

A tree T is a countable connected graph with a distinguished vertex called the root which has no cycles and such that each vertex belongs only to a finite number of edges. We use the symbol 0 for the root of T and we consider the tree as a directed graph where the edges go in the direction away from 0. Given a vertex σ in T there is a unique path $\pi(0, \sigma)$ connecting the root 0 to σ; in fact, we may identify σ and $\pi(0, \sigma)$. The number of edges of $\pi(0, \sigma)$ is called the *level number* of σ and indicated with $|\sigma|$; the root's level number is 0. For every vertex σ in T, except the root, the *parent* of σ is the unique vertex $\tau \in T$ with level number $|\sigma| - 1$ and with an edge to σ; conversely, σ is said to be a *child* of τ. We indicate the parent of σ with the symbol $\overleftarrow{\sigma}$. If two vertices have the same parent they are said to be *siblings*. We will consider trees with an infinite number of vertices and such that every vertex in the tree has the same number $k \geqslant 1$ of children; these are called *infinite k-ary trees* or, for short, *k-trees*.

Let E be the space of all infinite sequences $\varepsilon = (0 = \varepsilon_0, \varepsilon_1, \varepsilon_2, \dots)$ of vertices of a k-tree T with the property that, for all $i \geqslant 0$,

$$|\varepsilon_i| = i \quad \text{and} \quad \varepsilon_i = \overleftarrow{\varepsilon_{i+1}},$$

that is, ε_i is at level i and is the parent of ε_{i+1}. Think of ε as a path on the tree T connecting the root 0 to a limit point at an infinite level. The set E is called the *space of ends* of T. For $\varepsilon \in E$, set, for completeness, $\pi(0, \varepsilon) = \varepsilon$. Define $\bar{T} = T \cup E$.

For ξ and η in $\bar{T}$, $\xi \neq \eta$, define $\xi \wedge \eta$, the *confluent* of ξ and η, to be the vertex with highest level belonging both to $\pi(0, \xi)$ and $\pi(0, \eta)$; set $\xi \wedge \xi = \xi$. We now define a distance on $\bar{T}$ by setting, for every ξ and η in $\bar{T}$,

$$d(\xi, \eta) = \begin{cases} \exp(-|\xi \wedge \eta|), & \text{if } \xi \neq \eta, \\ 0, & \text{otherwise.} \end{cases} \tag{2.1}$$

After observing that, for all ξ, η and $\theta \in \bar{T}$,

$$|\xi \wedge \eta| \geqslant \min(|\xi \wedge \theta|, |\eta \wedge \theta|),$$

it is not difficult to prove that d is in fact a metric on $\bar{T}$ such that

$$d(\xi, \eta) \leqslant \max(d(\xi, \theta), d(\eta, \theta)),$$

for all $\xi, \eta, \theta \in \bar{T}$. The following result can be derived from Section 1.6 in ([2], pp. 219–221). A direct proof appears in Lemma 7.3 in Woess [11].

Proposition 2.1. *The space $\bar{T}$ with the metric d is compact and totally unconnected*; *T is a discrete subspace of $\bar{T}$ and E is a compact subspace of $\bar{T}$.*

Note that for $k \geqslant 1$ and $\varepsilon \in E$, the closed sphere $B_r(\varepsilon)$ with center ε and radius $r = \exp(-k)$ admits both representations:

$$B_r(\varepsilon) = \{\eta \in \bar{T} : d(\varepsilon, \eta) \leqslant r\} = \{\eta \in \bar{T} : d(\varepsilon, \eta) < \exp(1-k)\}.$$

Hence $B_r(\varepsilon)$ is simultaneously a closed and an open set in the topology induced by d on $\bar{T}$.

3. A reinforced dichotomous process indexed by a k-tree

We are now ready for the introduction of a stochastic process

$$X = \{X_\sigma : \sigma \in T\}$$

of Bernoulli random variables defined on a rich enough probability space $(\Omega, \mathscr{F}, P)$ and indexed by the vertices of a k-tree T. The process will also serve the purpose of introducing a more general process defined later in the paper.

The definition of the law of the process X is recursive on the levels of T. Let a and b be positive reals and set X_0 to be a Bernoulli random variable with parameter

$$p_0 = a/(a+b).$$

For $n \geqslant 0$, let $\mathscr{F}_n \subseteq \mathscr{F}$ be the sigma-field generated by the random variables X_σ with $|\sigma| \leqslant n$. Given $\mathscr{F}_n$, assume that the k^{n+1} random variables X_τ, with τ at level $n+1$, are conditionally independent and such that X_τ has Bernoulli distribution with parameter

$$p_\tau = \frac{a + \sum_{i=0}^{n} X_{\sigma_i}}{a+b+n+1},$$

if $\pi(0,\tau) = (0 = \sigma_0, \sigma_1, \ldots, \sigma_n, \tau)$. Therefore, random variables indexed by siblings have all the same conditional distribution.

We can form a mental picture of the process X by imagining that each vertex of the k-tree T is decorated with an urn containing balls of colors 1 and 0. In particular, at the root of the tree is sitting an urn containing a balls of color 1 and b balls of color 0; we sample a ball from this urn and we denote with X_0 the color of the ball extracted. A ball of color X_0 is then added to each urn corresponding to the k children of the root. Each of these urns initial composition is the same as the parent. That is, the composition of the urns of the children is *reinforced* according to the color extracted from the urn of their common parent. And so on; with the initial colors of each urn corresponding to those of its parent just before it is sampled, the composition of the urn corresponding to a vertex $\sigma \in T$ is the same as that of its parent plus an extra ball of color $X_{\overleftarrow{\sigma}}$ sampled from the urn of its parent $\overleftarrow{\sigma}$. The process X keeps track of the colors sampled from the urns corresponding to vertices of the tree T, while the process

$$p = \{p_\sigma : \sigma \in T\},$$

describes the proportion of balls of color 1 contained in the urns indexed by the vertices of the tree T. Observe that if $|\sigma| = n+1$, p_σ is $\mathscr{F}_n$-measurable while X_σ is measurable with respect to $\mathscr{F}_{n+1}$.

If $k = 1$, that is T is a 1-ary tree, the process X is a Pólya sequence, according to the definition of ([1], p. 353). For a general k-tree T, given any end $\varepsilon = (0 = \varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots) \in E$, the sequence of random variables $X_\varepsilon = (X_{\varepsilon_0}, X_{\varepsilon_1}, X_{\varepsilon_2}, \ldots)$ is a Pólya sequence. Hence, for $\varepsilon \in E$, X_ε is exchangeable; moreover, conditionally on the almost sure limit

$$p_\varepsilon = \lim_{n\to\infty} p_{\varepsilon_n}, \tag{3.1}$$

the random variables $X_{\varepsilon_0}, X_{\varepsilon_1}, X_{\varepsilon_2}, \ldots$ are independent and identically distributed with Bernoulli(p_ε) distributions. Finally, p_ε has a beta(a,b) distribution. Observe that p_ε is also the almost sure limit of the sequence of averages $\{n^{-1}\sum_{i=0}^{n-1} X_{\varepsilon_i}\}$.

On the other hand, given two different ends $\varepsilon = (0 = \varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots)$ and $\eta = (0 = \eta_0, \eta_1, \eta_2, \ldots)$ in E, the Pólya sequences X_ε and X_η have the same marginal law and are dependent. In fact, if $n = |\varepsilon \wedge \eta|$, then for $i \leqslant n$, we have $X_{\varepsilon_i} = X_{\eta_i}$ whereas, given $\mathscr{F}_n$, the subsequence $(X_{\varepsilon_{n+1}}, X_{\varepsilon_{n+2}}, \ldots)$ is conditionally independent from the subsequence $(X_{\eta_{n+1}}, X_{\eta_{n+2}}, \ldots)$ but with the same law. Whatever way we measure dependence, the intuitive result must be that the larger $n = |\varepsilon \wedge \eta|$ is, the greater the degree of dependence between X_ε and X_η. Or, in other words, the smaller the distance $d(\varepsilon, \eta)$ between ε and η, the greater the dependence between X_ε and X_η.

Before describing the process

$$\overrightarrow{p} = \{p_\varepsilon : \varepsilon \in E\},$$

in Subsection 3.1, we note an interesting property of the proportion of 1s at level n as $n \to \infty$. For $n \geqslant 0$, let us define

$$W_n = \frac{1}{k^n} \sum_{|\sigma|=n} p_\sigma,$$

to be the expected proportion of balls of color 1 at level $n+1$ conditioned on $\mathscr{F}_{n-1}$. Note that the random variable W_n is measurable with respect to $\mathscr{F}_{n-1}$.

Lemma 3.1. *The sequence $\{W_n\}$ is a bounded martingale with respect to the filtration $\{\mathscr{F}_{n-1}\}$; hence it converges almost surely to a random limit W.*

Proof. Trivially we have $W_n \in [0,1]$. In order to prove that $\{W_n\}$ is a martingale, for $n \geqslant 0$ we compute

$$\begin{aligned}
&E(W_{n+1}|\mathscr{F}_{n-1}) \\
&\quad = \frac{1}{k^{n+1}} \sum_{|\sigma|=n+1} E(p_\sigma|\mathscr{F}_{n-1}) \\
&\quad = \frac{1}{k^{n+1}} \sum_{|\sigma|=n+1} E\left\{ \frac{(a+b+n)p_{\overleftarrow{\sigma}} + X_{\overleftarrow{\sigma}}}{a+b+n+1} \middle| \mathscr{F}_{n-1} \right\} \\
&\quad = \frac{k}{k^{n+1}} \sum_{|\tau|=n} \frac{(a+b+n)p_\tau + E(X_\tau|\mathscr{F}_{n-1})}{a+b+n+1} \\
&\quad = \frac{1}{k^n} \sum_{|\tau|=n} p_\tau,
\end{aligned}$$

on a set of probability one. $\square$

The proportion of balls of color 1 generated by the urns associated with the vertices of the tree at level $n \geqslant 0$ is

$$Y_n = \frac{1}{k^n} \sum_{|\sigma|=n} X_\sigma.$$

The total number of balls of color 1 in the urns associated with vertices of the tree T at level $n+1$ is

$$\begin{aligned}
&k^{n+1}(a+b+n+1)W_{n+1} \\
&\quad = k * \text{total number of balls of color 1 in the urns at level } n \\
&\qquad + k * \text{number of balls of color 1 sampled from the urns at level } n \\
&\quad = k[k^n(a+b+n)W_n] + k \sum_{|\sigma|=n} X_\sigma \\
&\quad = k^{n+1}(a+b+n)W_n + k^{n+1} Y_n.
\end{aligned}$$

Hence,

$$(a+b+n+1)W_{n+1} = (a+b+n)W_n + Y_n$$

and thus

$$Y_n = (a + b + n)[W_{n+1} - W_n] + W_{n+1}.$$

Since $\{W_n\}$ is a martingale, this implies that

$$E(Y_n) = E(W_{n+1}) = W_0 = \frac{a}{a+b},$$

for all $n \geqslant 0$. For $k = 1$, $\{Y_n\}$ is a Pólya sequence of random variables valued 0 or 1; unless $a = 0$ or $b = 0$,

$$P(Y_n \neq Y_{n+1} \text{ for infinitely many } n) = 1$$

and thus $\{Y_n\}$ does not converge. However, let $W = \lim_{n\to\infty} W_n$.

Lemma 3.2. *For $k \geqslant 2$, $\lim_{n\to\infty} Y_n = W$ on a set of probability one.*

Proof. For $n \geqslant 0$, define

$$Z_n = Y_n - W_n = \frac{1}{k^n} \sum_{|\sigma|=n} (X_\sigma - p_\sigma).$$

For $n \geqslant 1$, Z_n is the average of k^n random variables bounded in absolute value by 1, with mean 0 and conditionally independent given $\mathscr{F}_{n-1}$. Hence, for $n \geqslant 1$,

$$\begin{aligned}
&E(Z_n^4)\\
&\quad = E\{E(Z_n^4|\mathscr{F}_{n-1})\}\\
&\quad = \frac{1}{k^{4n}} E\left[E\left\{\sum_{|\sigma|=n} (X_\sigma - p_\sigma)^4 + 3 \sum_{\sigma\neq\tau, |\sigma|=|\tau|=n} (X_\sigma - p_\sigma)^2 (X_\tau - p_\tau)^2 \middle| \mathscr{F}_{n-1}\right\}\right]\\
&\quad \leqslant \frac{k^n + 3k^n(k^n - 1)}{k^{4n}}\\
&\quad \leqslant \frac{3}{k^{2n}}.
\end{aligned}$$

Therefore, for $k \geqslant 2$,

$$E\left(\sum_{n=0}^{\infty} Z_n^4\right) = \sum_{n=0}^{\infty} E(Z_n^4) \leqslant \sum_{n=0}^{\infty} 3/k^{2n} < \infty.$$

Hence,

$$P\left(\sum_{n=0}^{\infty} Z_n^4 < \infty\right) = 1$$

and this implies that

$$P\left(\lim_{n\to\infty} Z_n = 0\right) = P\left(\lim_{n\to\infty} Z_n^4 = 0\right) = 1.$$

Since $W = \lim_{n\to\infty} W_n$ on a set of probability one, the proof of the lemma is complete. □

Example 3.3 (*Two-type reinforced branching process*). The process X can be considered as a reinforced branching process for modelling a population of organisms where two different types may be distinguished: types 0 and 1. Every individual of either type produces k offsprings, all of type $V \in \{0, 1\}$. If an individual's probability of producing offsprings of type 1 is $r/(r+s)$, then each of his offsprings has probability $(r+V)/(r+s+1)$ of producing offsprings of type 1. In this context, Y_n represents the fraction of type 1 individuals at generation $n+1$: the previous results asserts that this fraction converges to a random limit W as n grows to infinity. Note that, W being the limit of the martingale $\{W_n\}$,

$$E(W) = E\Big(\lim_{n\to\infty} Y_n\Big) = \frac{a}{a+b}.$$

3.1. The beta-blanket

We now focus on the process

$$\overrightarrow{p} = \{p_\varepsilon : \varepsilon \in E\},$$

of the limits of the averages of the process X along ends in E: these limits were defined in (3.1). We call the process $\overrightarrow{p}$ a *beta-blanket* with parameters (a, b). The law of $\overrightarrow{p}$ acts as a prior on the space of functions from the space of infinite sequences of 1s and 0s to the interval $[0, 1]$.

For every $\varepsilon \in E$, p_ε has a beta distribution with parameters (a, b). The joint distribution of p_ε and p_η, for $\varepsilon, \eta \in E$, is described by the next lemma as a mixture of products of beta distributions. Given any $c, d > 0$ and $x \in [0, 1]$, indicate with $B(c, d)$ the beta function evaluated in (c, d), that is

$$B(c,d) = \frac{\Gamma(c)\Gamma(d)}{\Gamma(c+d)},$$

with Γ the usual gamma function, and write $\Psi(x|c, d)$ for the cumulative distribution function of a beta distribution with parameters (c, d) evaluated at x.

Lemma 3.4. *Let* $\varepsilon, \eta \in E, \varepsilon \neq \eta$, *and set* $n = |\varepsilon \wedge \eta|$. *For all* $x_1, x_2 \in [0, 1]$,

$$P(p_\varepsilon \leqslant x_1, p_\eta \leqslant x_2) = \sum_{j=0}^{n+1} \binom{n+1}{j} \frac{B(a+j, b+n+1-j)}{B(a,b)} \prod_{i=1}^{2} \Psi(x_i|a+j, b+n+1-j).$$

Proof. Let $\varepsilon = (\varepsilon_0 = 0, \varepsilon_1, \varepsilon_2, \ldots), \eta = (\eta_0 = 0, \eta_1, \eta_2, \ldots) \in E$; then $\varepsilon_i = \eta_i$ for $i \leqslant n$. For $x_1, x_2 \in [0, 1]$,

$$
\begin{aligned}
&P(p_\varepsilon \leqslant x_1, p_\eta \leqslant x_2) \\
&\quad = E\{P(p_\varepsilon \leqslant x_1, p_\eta \leqslant x_2 | \mathscr{F}_n)\} \\
&\quad = E\{P(p_\varepsilon \leqslant x_1 | \mathscr{F}_n) P(p_\eta \leqslant x_2 | \mathscr{F}_n)\} \\
&\quad = E\left\{\prod_{i=1}^{2} \Psi\left(x_i \middle| a + \sum_{j=0}^{n} X_{\varepsilon_j}, b + n + 1 - \sum_{j=0}^{n} X_{\varepsilon_j}\right)\right\} \\
&\quad = E\left[E\left\{\prod_{i=1}^{2} \Psi\left(x_i \middle| a + \sum_{j=0}^{n} X_{\varepsilon_j}, b + n + 1 - \sum_{j=0}^{n} X_{\varepsilon_j}\right) \middle| p_\varepsilon\right\}\right] \\
&\quad = E\left\{\sum_{j=0}^{n+1} \binom{n+1}{j} p_\varepsilon^j (1 - p_\varepsilon)^{n+1-j} \prod_{i=1}^{2} \Psi(x_i | a + j, b + n + 1 - j)\right\} \\
&\quad = \sum_{j=0}^{n+1} \binom{n+1}{j} \frac{B(a+j, b+n+1-j)}{B(a,b)} \prod_{i=1}^{2} \Psi(x_i | a + j, b + n + 1 - j).
\end{aligned}
$$

The second equality holds because

$$(X_{\varepsilon_{n+1}}, X_{\varepsilon_{n+2}}, \ldots) \quad \text{and} \quad (X_{\eta_{n+1}}, X_{\eta_{n+2}}, \ldots)$$

are conditionally independent given $\mathscr{F}_n$; the next equality is true because the conditional distributions of p_ε and p_η, given $\mathscr{F}_n$, (the so called posterior distributions) are identical and equal to a beta distribution with parameters $(a + \sum_{j=0}^{n} X_{\varepsilon_j}, b + n + 1 - \sum_{j=0}^{n} X_{\varepsilon_j})$; the penultimate equality follows from the fact that, given p_ε, the random variables of the sequence X_ε are conditionally independent and identically distributed with Bernoulli(p_ε) distributions. Finally, the last equality holds because p_ε has a beta distribution with parameters (a, b). $\square$

Corollary 3.5. *For* $\varepsilon, \eta \in E, \varepsilon \neq \eta$,

(i) $$E\{(p_\varepsilon - p_\eta)^2\} = 2\, \frac{B(a+1, b+1)}{B(a,b)} \frac{1}{a + b + 1 - \log\, d(\varepsilon, \eta)},$$

hence the process $\overrightarrow{p}$ *is stochastically continuous.*

(ii) $$\mathrm{Corr}(p_\varepsilon, p_\eta) = \frac{1 - \log\, d(\varepsilon, \eta)}{a + b + 1 - \log\, d(\varepsilon, \eta)}.$$

Proof. Let $\varepsilon = (0 = \varepsilon_0, \varepsilon_1, \ldots)$ and $n = |\varepsilon \wedge \eta|$. Since we know the joint distribution of p_ε and p_η, for proving (i) we compute

$$\begin{aligned}
&E\{(p_\varepsilon - p_\eta)^2\} \\
&= 2\sum_{j=0}^{n+1} \binom{n+1}{j} \frac{B(a+j, b+n+1-j)}{B(a,b)} \operatorname{Var}[\mathrm{beta}(a+j, b+n+1-j)] \\
&= 2\sum_{j=0}^{n+1} \binom{n+1}{j} \frac{B(a+j, b+n+1-j)}{B(a,b)} \frac{(a+j)(b+n+1-j)}{(a+b+n+1)^2(a+b+n+2)} \\
&= \frac{2}{a+b+n+1} \frac{1}{B(a,b)} \sum_{j=0}^{n+1} \binom{n+1}{j} B(a+1+j, b+n+2-j) \\
&= \frac{2}{a+b+n+1} \frac{B(a+1, b+1)}{B(a,b)}.
\end{aligned}$$

Note that the first equality holds since, given $\sum_{i=0}^{n} X_{\varepsilon_i} = j$, p_ε and p_η are conditionally independent with distributions $\mathrm{beta}(a+j, b+n+1-j)$. Eq. (i) now follows, because $n = -\log d(\varepsilon, \eta)$. Therefore,

$$\lim_{d(\varepsilon,\eta)\to 0} E\{(p_\varepsilon - p_\eta)^2\} = 0,$$

which implies that, for every $\delta > 0$,

$$\lim_{d(\varepsilon,\eta)\to 0} P(|p_\varepsilon - p_\eta| > \delta) = 0$$

and hence $\vec{p}$ is stochastically continuous.

Now observe that

$$\operatorname{Corr}(p_\varepsilon, p_\eta) = 1 - \frac{E\{(p_\varepsilon - p_\eta)^2\}}{2\operatorname{Var}(p_\varepsilon)}, \tag{3.2}$$

since p_ε and p_η are identically distributed. But

$$\operatorname{Var}(p_\varepsilon) = \frac{ab}{(a+b)^2(a+b+1)}. \tag{3.3}$$

From (i) and Eqs. (3.2) and (3.3), (ii) follows easily. □

The correlation is an important feature of our prior. It simply states that the closer two ends are together, the higher the correlation between the corresponding probabilities.

The next lemma gives a representation for the conditional distribution of p_η, given p_ε, as a mixture of beta distributions with binomial weights.

Lemma 3.6. *Let* $\varepsilon, \eta \in E, \varepsilon \neq \eta$, *and set* $n = |\varepsilon \wedge \eta|$. *For* $x \in [0,1]$, *on a set of probability one*,

$$P(p_\eta \leqslant x | p_\varepsilon) = \sum_{j=0}^{n+1} \binom{n+1}{j} p_\varepsilon^j (1-p_\varepsilon)^{(n+1-j)} \Psi(x | a+j, b+n+1-j).$$

Proof. Let $\varepsilon = (\varepsilon_0 = 0, \varepsilon_1, \varepsilon_2, \ldots)$. Given $\mathscr{F}_n$, the variables p_ε and p_η are conditionally independent having beta distribution with parameters

$$\left(a + \sum_{j=0}^{n} X_{\varepsilon_j}, b+n+1-\sum_{j=0}^{n} X_{\varepsilon_j}\right).$$

Hence, for $x \in [0,1]$,

$$\begin{aligned}
&P(p_\eta \leqslant x | p_\varepsilon)\\
&\quad = E\{P(p_\eta \leqslant x | p_\varepsilon, \mathscr{F}_n) | p_\varepsilon\}\\
&\quad = E\{P(p_\eta \leqslant x | \mathscr{F}_n) | p_\varepsilon\}\\
&\quad = E\left\{\Psi\left(x \,\middle|\, a+\sum_{j=0}^{n} X_{\varepsilon_j}, b+n+1-\sum_{j=0}^{n} X_{\varepsilon_j}\right) \,\middle|\, p_\varepsilon\right\}\\
&\quad = \sum_{j=0}^{n+1} \binom{n+1}{j} p_\varepsilon^j (1-p_\varepsilon)^{n+1-j} \Psi(x | a+j, b+n+1-j),
\end{aligned}$$

on a set of probability one, where the last equality follows from the fact that, given p_ε, the random variables of the sequence X_ε are conditionally independent and identically distributed with distribution Bernoulli(p_ε). $\square$

Corollary 3.7. *Let* $\varepsilon, \eta \in E, \varepsilon \neq \eta$; *then*

$$E(p_\eta | p_\varepsilon) = \frac{a + (|\varepsilon \wedge \eta| + 1) p_\varepsilon}{a + b + |\varepsilon \wedge \eta| + 1}.$$

Given p_ε, the variables of the sequence $X_\varepsilon = (X_{\varepsilon_0} = X_0, X_{\varepsilon_1}, \ldots)$ are conditionally independent and identically distributed with distribution Bernoulli(p_ε). For application in predictive inference, it is interesting to compute the conditional distribution of the random variables of the sequence $X_\eta = (X_{\eta_0} = X_0, X_{\eta_1}, \ldots)$ given p_ε, for $\varepsilon \neq \eta$.

Lemma 3.8. *Let* $\varepsilon, \eta \in E, \varepsilon \neq \eta$, *and set* $n = |\varepsilon \wedge \eta|$. *For* $s \geqslant 0$ *and* $i_0, \ldots, i_s \in \{0,1\}$,

$$\begin{aligned}
&P(X_{\eta_0} = i_0, \ldots, X_{\eta_s} = i_s | p_\varepsilon)\\
&\quad = \begin{cases} p_\varepsilon^{\xi_s} (1-p_\varepsilon)^{s+1-\xi_s}, & \text{if } s \leqslant n,\\[2ex] p_\varepsilon^{\xi_n} (1-p_\varepsilon)^{n+1-\xi_n} \dfrac{B(a+\xi_s, b+s+1-\xi_s)}{B(a+\xi_n, b+n+1-\xi_s)}, & \text{if } s > n, \end{cases}
\end{aligned}$$

where $\xi_s = \sum_{r=0}^{s} i_r$.

Proof. Given p_ε, for $s \leqslant n$ the random variables $X_{\eta_0}, \ldots, X_{\eta_s}$ are conditionally independent with distribution Bernoulli(p_ε). Hence, take $s > n$. Then,

$$\begin{aligned}
&P(X_{\eta_0} = i_0, \ldots, X_{\eta_s} = i_s | p_\varepsilon) \\
&\quad = E\{P(X_{\eta_0} = i_0, \ldots, X_{\eta_s} = i_s | \mathscr{F}_n, p_\varepsilon) | p_\varepsilon\} \\
&\quad = E\{I(X_{\eta_0} = i_0, \ldots, X_{\eta_n} = i_n) P(X_{\eta_{n+1}} = i_{n+1}, \ldots, X_{\eta_s} = i_s | \mathscr{F}_n, p_\varepsilon) | p_\varepsilon\} \\
&\quad = E\{I(X_{\eta_0} = i_0, \ldots, X_{\eta_n} = i_n) P(X_{\eta_{n+1}} = i_{n+1}, \ldots, X_{\eta_s} = i_s | \mathscr{F}_n) | p_\varepsilon\} \\
&\quad = \frac{B\left(a + \sum_{r=0}^{s} i_r, b + s + 1 - \sum_{r=0}^{s} i_r\right)}{B\left(a + \sum_{r=0}^{n} i_r, b + n + 1 - \sum_{r=0}^{s} i_r\right)} P(X_{\eta_0} = i_0, \ldots, X_{\eta_n} = i_n | p_\varepsilon) \\
&\quad = \frac{B\left(a + \sum_{r=0}^{s} i_r, b + s + 1 - \sum_{r=0}^{s} i_r\right)}{B\left(a + \sum_{r=0}^{n} i_r, b + n + 1 - \sum_{r=0}^{s} i_r\right)} p_\varepsilon^{\sum_{r=0}^{s} i_r} (1 - p_\varepsilon)^{s+1-\sum_{r=0}^{s} i_r}.
\end{aligned}$$

Note that the third equality holds because p_ε and the random variables $X_{\eta_{n+1}}, \ldots, X_{\eta_s}$ are conditionally independent, given $\mathscr{F}_n$. $\square$

Two pictures of beta blankets are presented in Figs. 1 and 2. These were generated to level 16 and so are approximate. Of course we cannot compute numerically an exact beta blanket. Fig. 1 is with $a = b = \frac{1}{2}$ and Fig. 2 is with $a = b = 5$.

4. Illustration: binary regression with balanced design

The prior described in Section 3 is fundamentally different from the prior of Diaconis and Freedman [5]. Our prior for the n balanced design model is the law of $\{p_\sigma : |\sigma| = n\}$, which is a probability on the space of functions from the set of $n+1$ sequences of 1s and 0s to the interval [0, 1]. The posterior in this case is easy to find. Let us denote by $\mathscr{X}_n = \{X_\sigma : |\sigma| < n\}$, for $n > 0$. It is quite clear that the data form the set

$$X^n = \{X_\sigma : |\sigma| = n\}.$$

Hence, of interest is the posterior distribution of $\mathscr{X}_n$ given X^n. While this is difficult to manage mathematically, it is quite straightforward to sample using a Gibbs

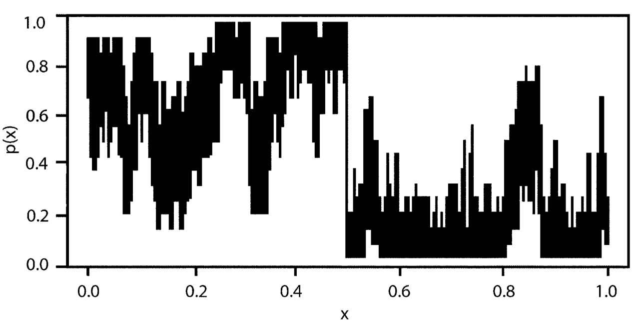

Fig. 1. Simulated beta blanket with $a = b = \frac{1}{2}$.

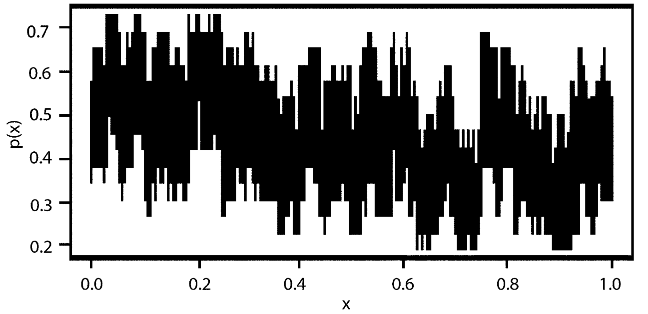

Fig. 2. Simulated beta blanket with $a = b = 5$.

sampler; see, for example, Smith and Roberts [9]. This is because, for any $X_\sigma \in \mathscr{X}_n$, we have access to the conditional distribution of X_σ given its ancestors and descendants, to level n. Then $p(X_\sigma | \cdots)$ can be sampled by noting that

$$p(X_\sigma = 1 | \cdots) \propto p(X_\sigma = 1 | \text{ancestors}) \times p(\text{descendents} | X_\sigma = 1).$$

Consequently, a Gibbs sampler is easy to implement. Of course, a Bayes estimate of $\overrightarrow{p}$ is then available from the output of the Gibbs sampler.

It is also possible to deal with an unbalanced design. The likelihood function is easy to write down and, as with the balanced design, it is possible to find posterior summaries via a Gibbs sampler.

5. A Dirichlet reinforced process indexed by a k-tree

The construction of the process introduced in Section 3 can be mimicked in a more general setting for defining a reinforced stochastic process $X = \{X_\sigma : \sigma \in T\}$ of random variables indexed by the vertices of a k-tree T and with values in a Polish space S endowed with its Borel sigma-field $\mathscr{S}$. As before, we assume that the random variables of the process X are defined on a rich enough probability space $(\Omega, \mathscr{F}, P)$ and we specify the law of X recursively on the levels of the tree T.

Let G_0 be a probability distribution on S and $c > 0$ a constant. Set X_0 to be a random variable with values in S and probability distribution G_0. For $n \geqslant 0$ let $\mathscr{F}_n \subseteq \mathscr{F}$ be the sigma-field generated by the random variables X_σ with $|\sigma| \leqslant n$. Given $\mathscr{F}_n$, assume that the k^{n+1} random variables X_τ, with τ at level $n+1$, are conditionally independent and such that X_τ has values in S and probability distribution

$$G_\tau = \frac{cG_0 + \sum_{i=0}^{n} \delta_{X_{\sigma_i}}}{c+n+1}, \tag{5.1}$$

if $\pi(0,\tau) = (0 = \sigma_0, \sigma_1, \ldots, \sigma_n, \tau)$; for $s \in S$, δ_s indicates the point mass at s. We recover the dichotomous reinforced process of the previous section if we set $S = \Re$,

$$G_0 = \frac{b}{a+b}\,\delta_0 + \frac{a}{a+b}\,\delta_1$$

and $c = a + b$. In point of fact, if I is the symbol for the indicator function, for every $B \in \mathscr{S}$, the process $\{I[X_\sigma \in B] : \sigma \in T\}$ is a dichotomous reinforced process indexed by the k-tree T.

Given an end $\varepsilon = (0 = \varepsilon_0, \varepsilon_1, \ldots) \in E$, the sequence of random variables $X_\varepsilon = (X_{\varepsilon_0}, X_{\varepsilon_1}, X_{\varepsilon_2}, \ldots)$ is a Pólya sequence with parameter cG_0. That is, the initial color X_0, which are assumed to exist on a continuum, is drawn from G_0. The next color X_{ε_1} is drawn from

$$\frac{cG_0 + \delta_{X_0}}{c+1}$$

and in general $X_{\varepsilon_{n+1}}$ is drawn from

$$\frac{cG_0 + \sum_{i=0}^{n} \delta_{X_{\varepsilon_i}}}{c+n+1}.$$

See Blackwell and MacQueen [1], where it is proved that X_ε is exchangeable and that the random distribution functions of the sequence $\{G_{\varepsilon_n}\}$ weakly converge to a Dirchlet process G_ε with parameter cG_0 on a set of probability one. Moreover, given G_ε, the random variables of the sequence X_ε are conditionally independent and identically distributed with distribution G_ε.

In order to obtain a result analogous to Lemma 3.2 in this more general setting, for $n \geqslant 0$, define G_n to be the empirical distribution function of the random variables X_σ with $|\sigma| = n$, that is

$$G_n(B) = \frac{1}{k^n} \sum_{\{\sigma : |\sigma| = n\}} I[X_\sigma \in B],$$

for every $B \in \mathscr{S}$. Here we assume that $k \geqslant 2$. Given $B \in \mathscr{S}$, the sequence of real random variables $\{G_n(B)\}$ converges with probability one because of Lemma 3.2. For every finite measurable partition $(B_1, \ldots, B_r)$ of S, define

$$(G(B_1), \ldots, G(B_r)) = \left(\lim_{n\to\infty} G_n(B_1), \ldots, \lim_{n\to\infty} G_n(B_r) \right). \tag{5.2}$$

Let $\mathscr{P}$ be the class of probability measures defined on the Borel σ-field $\mathscr{S}$ of S; endow $\mathscr{P}$ with the topology of weak convergence and write $\sigma(\mathscr{P})$ for the Borel σ-field in $\mathscr{P}$. With these assumptions $\mathscr{P}$ becomes a separable and complete metric space.

Theorem 5.1. *Eqs.* (5.2) *define a random element G of $\mathscr{P}$. Furthermore, G is the limit in distribution of the sequence* $\{G_n\}$.

Proof. For proving the first part of the theorem, we may check the consistency conditions (C_1)–(C_4) of Cifarelli et al. [3]. See also Regazzini and Petris [8].

For proving the second part, it is enough to show that the sequence of measures induced on $(\mathscr{P}, \sigma(\mathscr{P}))$ by the random elements $G_n \in \mathscr{P}$, $n = 1, 2, \ldots$, is tight. Given $d > 0$, let K_t, $t = 1, 2, \ldots$, be a compact set of S such that $G_0(K_t^c) \leqslant d/t^3$ and define

$$M_t = \{H \in \mathscr{P} : H(K_t^c) \leqslant 1/t\}.$$

The set $M = \bigcap_{t=1}^{\infty} M_t$ is compact in $\mathscr{P}$. For $n = 1, 2, \ldots$ and $t = 1, 2, \ldots$, the sequence $\{G_n(K_t^c)\}$ is a martingale because of Lemma 3.1, thus $E\{G_n(K_t^c)\} = G_0(K_t^c)$.

Hence

$$P\{G_n(K_t^c) > 1/t\} \leqslant tG_0(K_t^c) \leqslant d/t^2.$$

Therefore, for every $n = 1, 2, \ldots,$

$$P(G_n \in M) \geqslant 1 - \sum_{t=1}^{\infty} P\{G_n(K_t^c) > 1/t\} \geqslant 1 - d \sum_{t=1}^{\infty} 1/t^2$$

and this proves that the sequence of measures induced on $(\mathscr{P}, \sigma(\mathscr{P}))$ by the empirical distribution functions G_n, $n = 1, 2, \ldots,$ is tight. $\square$

Example 5.2 (*Colonization of a region*). Let $S = \Re^3$, and consider a population of organisms living in S: at every time period $n = 0, 1, 2, \ldots,$ each individual of the population chooses a location in S, produces $k \geqslant 2$ offsprings and then dies. At time $n = 0$ there is only one individual, the first colonist, it chooses its location at random according to the distribution G_0, produces k offsprings and dies. At time $n = 1$, before reproduction, each of the k offsprings chooses its own location in S at random according to a distribution that is the same as that of the parent but for the fact that the location chosen by the parent has been reinforced with a unitary weight: how strongly this weight modifies the parent's distribution depends on c, higher values of c implying weaker effects. Precisely, the locations for each of the k individuals forming the first generation are independently generated by the distribution

$$\frac{cG_0 + \delta_{X_0}}{c + 1}.$$

For instance, we may think that c quantifies a cultural drive generically called the force of ancestral tradition; the higher is c, the less the parent's location influences the distribution generating locations for its offsprings. And so on forever, with the force of ancestral tradition growing of one unit generation after generation: hence, individual τ of the $(n+1)$th generation will choose its location at random, and independently from the other individuals of its generation, according to the distribution

$$G_\tau = \frac{(c+n)G_{\overleftarrow{\tau}} + \delta_{X_{\overleftarrow{\tau}}}}{c+n+1},$$

$\overleftarrow{\tau}$ being τ's parent.

The previous theorem affirms that the distribution on S of the nth generation is random but converges, as n grows to infinity, to the random distribution G. Note that if the support of G_0 is a proper subset R of S, the support of G is contained in R with probability one.

5.1. The Dirichlet-blanket

Analogously to Section 2.1, we now study the process

$$\overrightarrow{G} = \{G_\varepsilon : \varepsilon \in E\}.$$

For every $\varepsilon \in E$, we know that G_ε is a random element of $\mathscr{P}$ whose law is that of a Dirichlet process with parameter cG_0. We call $\overrightarrow{G}$ a *Dirchlet-blanket* with parameter cG_0. Note that, for every $B \in \mathscr{S}$,

$$\overrightarrow{G}(B) = \{G_\varepsilon(B) : \varepsilon \in E\}$$

is a beta-blanket with parameters $(cG_0(B), cG_0(B^c))$; hence the results of Section 3.1 apply.

The analogy of Lemma 3.4 would illustrate the joint probability law of G_ε and G_η for $\varepsilon, \eta \in E, \varepsilon \neq \eta$. In fact, the bivariate process (G_ε, G_η) has been studied by Walker and Muliere [10] under the name of *bivariate Dirichlet process*.

Let $c > 0$, r a non-negative integer, G_0 a probability distribution on S; let F_1 and F_2 be two random elements of $\mathscr{P}$ and N a point process on the real line. Assume that F_1 is a Dirichlet process with parameter cG_0 and that, given F_1, for every finite measurable partition $(B_1, \ldots, B_h)$ of S, the conditional distribution of $(N(B_1), \ldots, N(B_h))$ is multinomial with parameters $(r, (F_1(B_1), \ldots, F_1(B_h)))$. Finally, given N, assume that the conditional law of F_2 is that of a Dirichlet process with parameter $(cG_0 + N)$. Then, Walker and Muliere [10] call the process (F_1, F_2) a bivariate Dirichlet process with parameters (cG_0, r).

Theorem 5.3. *For* $\varepsilon, \eta \in E$, $\varepsilon \neq \eta$, (G_ε, G_η) *is a bivariate Dirichlet process with parameters* $(cG_0, 1 + |\varepsilon \wedge \eta|)$.

Proof. Let $\varepsilon = (0 = \varepsilon_0, \varepsilon_1, \varepsilon_2, \ldots)$ and $n = |\varepsilon \wedge \eta|$. Given $\mathscr{F}_n$, G_ε and G_η are two conditionally independent Dirichlet processes both with parameter $(cG_0 + \sum_{i=0}^n \delta_{X_{\varepsilon_i}})$. Furthermore, given G_ε, the random variables of the Pólya sequence $X_\varepsilon = (X_{\varepsilon_0}, X_{\varepsilon_1}, X_{\varepsilon_2}, \ldots)$ are conditionally independent and identically distributed with distribution G_ε; therefore, for every finite measurable partition $(B_1, \ldots, B_h)$ of S, the conditional distribution of

$$\left(\sum_{i=0}^n \delta_{X_{\varepsilon_i}}(B_1), \ldots, \sum_{i=0}^n \delta_{X_{\varepsilon_i}}(B_h)\right)$$

is multinomial with parameters $\{n+1, (G_\varepsilon(B_1), \ldots, G_\varepsilon(B_h))\}$. The result follows. $\square$

For $\varepsilon, \eta \in E$, the dependence between the processes G_ε and G_η is such that, for every $B \in \mathscr{S}$,

$$\mathrm{Corr}(G_\varepsilon(B), G_\eta(B)) = \frac{1 - \log d(\varepsilon, \eta)}{c + 1 - \log d(\varepsilon, \eta)}.$$

This follows from Corollary 3.5, as well as from the fact that, for $\varepsilon \neq \eta$, (G_ε, G_η) is a bivariate Dirichlet process. In fact, it was the need to express this type of dependence in a Bayesian nonparametric context, that motivated the introduction of the bivariate Dirichlet process of Walker and Muliere. Along this path, for $r \geqslant 2$ and $\eta^1, \ldots, \eta^r$ different ends of E, we may say that $(G_{\eta^1}, \ldots, G_{\eta^r})$ is an r-variate Dirichlet process after observing that, for $B \in \mathscr{S}$,

$$\mathrm{Corr}(G_{\eta^i}(B), G_{\eta^j}(B)) = \frac{1 - \log\, d(\eta^i, \eta^j)}{c + 1 - \log\, d(\eta^i, \eta^j)}, \tag{5.3}$$

for $i,j \in \{1, \ldots, r\}$. Note that if $r \leqslant k$, we may take $\eta^1, \ldots, \eta^r$ such that $|\eta^i \wedge \eta^j|$ is the same for every $i,j \in \{1, \ldots, r\}$. Hence, the correlation (5.3) is the same for all couples of components of the multivariate process $(G_{\eta^1}, \ldots, G_{\eta^r})$.

If we assume data arises as the X values from level n, then Bayesian nonparametric inference can be accomplished in a similar way as outlined in Section 3. The difference being that the conditional distribution of X_σ given its children and ancestors is based on the mixture distributions appearing in (5.1).

6. General partially exchangeable processes

A generalization of previous sections, which we will describe briefly, is based on a prior distribution on the space of density functions. Let this prior be denoted by Π. The random variable X_0 is distributed according to the prior predictive density $f_0(x) = \int f(x)\, \Pi(\mathrm{d}f)$. For $n \geqslant 0$, conditionally on the random variables X_σ with $|\sigma| \leqslant n$, the random variables X_τ with τ at level $n+1$ are independent and distributed according to the predictive density

$$f_\sigma(x) = \int f(x) \Pi_\tau(\mathrm{d}f),$$

where

$$\Pi_\tau(\mathrm{d}f) = \frac{\prod_{i=0}^{n} f(X_{\sigma_i})\, \Pi(\mathrm{d}f)}{\int \prod_{i=0}^{n} f(X_{\sigma_i})\, \Pi(\mathrm{d}f)}$$

and $\pi(0, \tau) = (0 = \sigma_0, \sigma_1, \ldots, \sigma_n, \tau)$. It is clear that X_ε is exchangeable for each $\varepsilon \in E$ and the empirical distribution function of the sequence converges on a set of probability one to a random distribution function chosen from Π.

The benefit when the prior is supported by densities (which is not the case with the Dirichlet process) is that the random distribution functions from Π will have densities. In this case the random sequence of predictive densities also converges, in the Hellinger sense, on a set of probability one to a random density function chosen from the prior Π. Hence we can construct, analogously to Sections 3 and 5, a Π-blanket of random density functions. The partially exchangeable process X on which this blanket is constructed will also be useful for regression problems.

7. Discussion

We have constructed processes which are formed via dependent exchangeable sequences. The collection of these sequences would be termed *partially exchangeable* in the language of Finetti [4]. It is well known that partially exchangeable sequences have applications in Bayesian regression problems.

The paper can be seen as a natural extension of previous work undertaken by the authors. One path of the tree indexes a sequence of exchangeable variables whose law is constructed via reinforcement; the branching allows for the extension to a partially exchangeable process where the dependence among exchangeable subsequences corresponding to different paths is described by the geometry of the tree; the closer the paths on the tree, that is the closer the covariates associated with the exchangeable sequences, the higher the dependence. Along the same idea one could define a process generated by dependent subsequences of random variables, each subsequence corresponding to a path on a tree having a law constructed via reinforcement; as an example, consider the situation where each path of the tree indexes a recurrent reinforced urn process (Muliere et al. [6]), that is, a mixture of Markov chains.

Further work in this area would involve the possibility of more complicated tree structures to cater for more complicated regression structures and also the possibility of randomly generated tree structures.

References

[1] D. Blackwell, J.B. MacQueen, Ferguson distributions via Pólya-urn schemes, Ann. Statist. 2 (1973) 353–355.

[2] P. Cartier, Functiones harmoniques sur un arbre, in: Symposia Mathematica, vol. 9, Academic Press, London, 1972, pp. 203–270.

[3] D.M. Cifarelli, P. Muliere, P. Secchi, Prior processes for Bayesian nonparametrics, Technical Report, 377/P, Politecnico di Milano, 1999.

[4] B. de Finetti, Sur la condition d'equivalence partielle, VI Colloque Geneve, Acta. Sci. Ind. Paris (1938) 739.

[5] P. Diaconis, D.A. Freedman, Nonparametric binary regression: a Bayesian approach, Ann. Statist. 21 (1993) 2108–2137.

[6] P. Muliere, P. Secchi, S.G. Walker, Urn schemes and reinforced random walks, Stochastic Process. Appl. 88 (2000) 59–78.

[7] P. Muliere, P. Secchi, S.G. Walker, Reinforced random processes in continuous time, Stochastic Process. Appl. 104 (2003) 117–130.

[8] E. Regazzini, G. Petris, Some critical aspects on the use of exchangeability in Statistics, J. Ital. Statist. Soc. 1 (1992) 103–130.

[9] A.F.M. Smith, G.O. Roberts, Bayesian computations via Gibbs sampler and related Markov chain Monte Carlo methods, J. Roy. Statist. Soc. Ser. B 55 (1993) 3–23.

[10] S.G. Walker, P. Muliere, A bivariate dirichlet process, Statist. Probab. Lett. 64 (2003) 1–7.

[11] W. Woess, Catene di Markov e Teoria del Potenziale nel Discreto, Quaderni dell'Unione Matematica Italiana, vol. 41, Pitagora Editrice, Bologna, 1996.